Rev. Roum. Sci. Techn.– Électrotechn. et Énerg.
Vol. **70**,*1*, pp. 103–108, Bucarest, 2025

CLUSTER AND LINEAR REGRESSION ANALYSIS OF CHINESE AND MALAYSIAN RESEARCH COLLABORATION BASED ON BIG DATA AND SCIBERT

GUANSU WANG[1], ZHIHONG HUANG[2], SAMEER KUMAR[*]

Keywords: Research collaboration; Cluster analysis; Linear regression; “Belt and road” initiative.

This study employs big data analytics and SciBERT to conduct cluster and linear regression analyses of research collaboration between China and Malaysia. Originating from China's “Belt and Road” initiative, this collaboration has evolved into a comprehensive strategic partnership, fostering advancements in economics, agriculture, technology, education, and more. Establishing multi-dimensional strategic relationships aligns with global trends, emphasizing technological innovation's pivotal role. Both nations have implemented policies to boost science and technology, influencing their collaborative efforts. Research collaboration serves as a driving force for technological progress, intertwining with cultural exchange. The study focuses on the trends, characteristics, and influencing factors of China-Malaysia research collaboration using data from the Web of Science. The findings provide insights for optimizing collaboration models and guiding future policies, contributing to the communication and development between China and Malaysia.

1. INTRODUCTION

China introduced the “Belt and Road” initiative in 2013, establishing cooperative frameworks highlighting its proactive global role. As a strategic partner of ASEAN, China aims to foster collective progress with neighboring countries [1]. Malaysia, a pivotal Southeast Asian participant in the initiative, contributes to policy coordination, trade facilitation, infrastructure, finance, and cultural exchange [2]. Since establishing diplomatic relations in 1974, China and Malaysia have maintained strong ties, culminating in the 2023 China-Malaysia Community of Shared Future. Their cooperation spans economics, agriculture, technology, education, culture, medicine, and defense [3]. Research collaboration has become a key aspect of their bilateral relations, intertwined with cultural exchanges. Over the years, both countries have introduced policies to enhance academic exchanges, talent development, and joint scientific projects.

This study leverages Web of Science data to analyze the patterns, characteristics, and influencing factors of the China-Malaysia research collaboration. Using cluster and linear regression analysis, it aims to provide actionable insights and policy recommendations for enhancing future cooperation.

2. LITERATURE REVIEW

Research on China-Malaysia collaborative research often focuses on technological cooperation within the “Belt and Road” initiative or in the broader context of ASEAN collaboration.

Firstly, in terms of collaboration trends, scholars analyzing the collaborative research in the “Belt and Road” region and countries along it have found that ASEAN countries have the highest share of international collaborative papers, approximately 48.7%. The joint research output between China and ASEAN countries is significantly ahead of other “Belt and Road” countries [4,5]. China occupies a central position in the international collaborative research network of the “Belt and Road”, producing many collaborative research papers with Singapore, Russia, and India. At the same time, collaborations with Malaysia are relatively fewer. From another perspective, China's “Belt and Road” collaboration intensity (< 30%) is significantly lower than that of other countries, indicating substantial space and opportunities for improvement in China's research collaboration with countries along the “Belt and Road” [4]. Kumar & Jan suggested a significant potential for Malaysia to enhance its research collaboration with countries along the “Belt and road” [6].

Furthermore, in terms of collaborative areas, an analysis by Ye et al. on the characteristics of China's research collaboration with "Belt and Road" countries revealed that Singapore and Malaysia are the leading collaborators with China, particularly in patent numbers [7]. The primary collaborative areas encompass semiconductor devices, digital data processing, wireless communication networks, and chemical or physical methods. In a study analyzing China's research collaboration with ASEAN, Chen and Xu found that collaborations between countries primarily occur in advantageous and distinctive fields, focusing predominantly on science and engineering. In contrast, collaborations between the humanities and social sciences are relatively limited [8]. Malaysian universities collaborate extensively internationally, particularly in physics and astronomy, chemistry, agriculture and biological sciences, engineering, health professions, and computer science [9]. Although different subjects may demonstrate varying collaboration trends, a general overview suggests a growing trend of collaborative research across subjects. The number of co-authored papers, particularly those with international co-authorship, has consistently increased [10].

Finally, regarding the factors influencing research collaboration, a study by Davidson Frame and Carpenter concluded that both science-related factors (national R&D efforts and research enterprise) and non-science factors (geography, language, culture, and politics) determine the extent and patterns of international collaborative behavior [11]. An analysis of research collaboration between China and ASEAN revealed that the overall financial, research, communication, and international exchange positively correlates with the likelihood of cooperation [8]. Additionally, collaboration is influenced by factors such as national distance, social distance, and disciplinary similarity.

[1] Asia-Europe Institute, Universiti Malaya, Kuala Lumpur, Malaysia. E-mail: s2128778@siswa.um.edu.my
[2] Zhuhai College of Jilin University, Zhuhai, Guangdong, China. E-mail: zayvion_huang@outlook.com
[*] Corresponding Author: Asia-Europe Institute, Universiti Malaya, Kuala Lumpur, Malaysia. E mail: sameer@um.edu.my

DOI: 10.59277/RRST-EE.2025.1.18

Notably, the global research collaboration network demonstrates a trend of solid alliances, where countries with similar levels of economic development exhibit increased chances and scale of research collaboration [12].

In summary, existing research primarily analyzes collaboration between China and "Belt and Road" or ASEAN countries, with few focusing specifically on China-Malaysia trends. Comprehensive studies on collaboration patterns and influencing factors are lacking, limiting their policy guidance. This study addresses this gap by analyzing 37 years (1988–2024) of Web of Science data to examine China-Malaysia research trends, identify influencing factors, and provide policy recommendations to enhance collaboration.

3. DATA AND METHODOLOGY

Academic publication data reflect the state of collaboration between Chinese and Malaysian scholars, making it a valuable subject in cross-cultural and bibliometric research [6]. Cluster analysis, commonly used in bibliometric studies, identifies connections and categorizes data to uncover research trends and topics [13,14]. Linear regression examines influencing factors. This study applies cluster analysis to text data from co-authored publications. It uses linear regression on numeric data to explore research trends, key fields, and factors influencing China-Malaysia academic collaboration, offering recommendations to enhance cross-country cooperation.

3.1. DATA COLLECTION & CLEANING

The Web of Science Core Collection (WoSCC), a widely used bibliometric database, enables data retrieval via queries like Topic (TS), Title (TI), and Country/Region (CU). Using the query 'CU=China* AND CU=Malaysia*', 21,411 co-authored publications were retrieved on Dec. 13, 2023. After excluding records with missing key information, 21,409 publications remained. This dataset includes titles, abstracts, research areas, authors' full names, and affiliations. Research areas were classified into five categories—Arts & Humanities, Life Science & Biomedicine, Physical Sciences, Social Sciences, and Technology—based on the Web of Science framework [15].

To prepare textual data (titles and abstracts) for Natural Language Processing (NLP), NLTK, a Python library, was used. Titles were cleaned of punctuation to facilitate vectorization using a BERT model. Abstracts were tokenized, stopwords removed, and stemmed with NLTK for word frequency analysis, improving machine recognition and analysis accuracy [16].

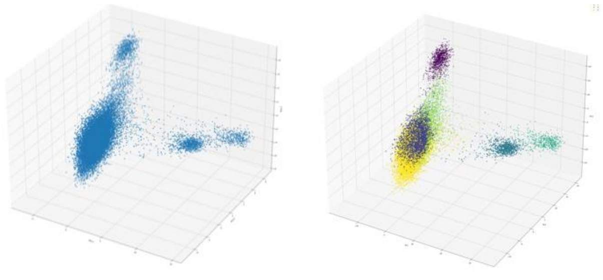
3.2. CLUSTERING

Clustering algorithms divide datasets into groups based on criteria like Euclidean distance in multidimensional space. The goal is to maximize similarity within clusters and minimize similarity between clusters, ensuring data points in the same cluster are highly similar, and those in different clusters are distinct.

For textual data in natural language, preprocessing transforms the data into multidimensional vectors [17], using methods such as TF-IDF, Word2Vec [18,19], or advanced Large Language Models (LLMs) like BERT and GPT-3.5 [20,21]. After vectorization, an appropriate clustering algorithm and the optimal number of clusters (K) are selected, followed by result evaluation.

BERT, introduced by Google, is a large language model that, compared to TF-IDF and Word2Vec, can process textual data at a contextual level, directly performing the vectors of sentences [22]. SciBERT, a pre-trained BERT model trained on scientific text databases [23], is specifically developed for scientific text comprehension and is highly used in papers that process scientific textual data [24–26], which aligns with the demand of this study. Using SciBERT, the title data were vectorized with 768 dimensions. The number of clusters (K) was determined using the elbow method and SSE (Sum Squared Errors), which showed that in the range of 4–6 was appropriate [27].

The 768-dimensional data was projected to three-dimensional spaces for visualization purposes (Fig. 1a). The visual results indicated that the optimal K was around 4. According to the data's distribution status and the high-dimensional characteristics, the Spectral Clustering algorithm was chosen for clustering.



(a) (b)

Fig. 1 – Three-dimensional projection, K=1 (a), K=6 (b).

Spectral Clustering was utilized, and K was selected as 4 first. Based on the results (K=4), one of the clusters was observed to have significantly more entries compared to the others. Considering the data's real-world meaning, this disproportion seemed unreasonable. Therefore, after several attempts, the results obtained when K=6 provided the best explanation (see Fig. 1b) for the current state of academic collaboration between Chinese and Malaysian scholars. Hence, this clustering result was selected for subsequent analysis in this study.

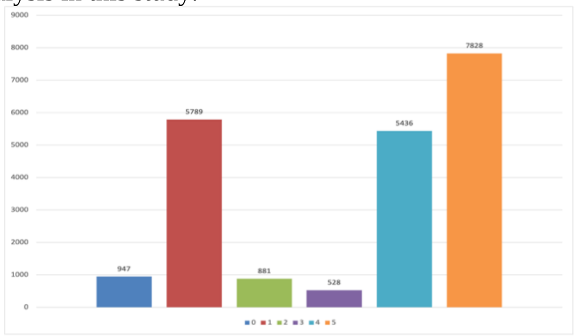


Fig. 2 – Data count.

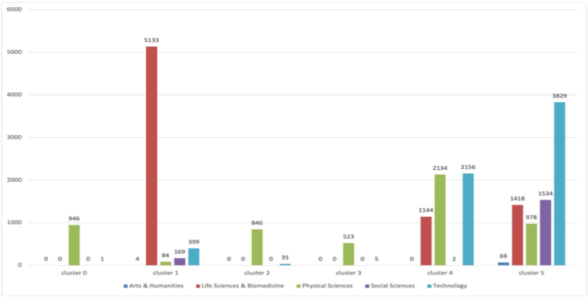


Fig. 3 – Research areas count.

3.3. LINEAR REGRESSION

After gaining a basic understanding of the collaborative trends between China and Malaysia, it is essential to investigate further the factors influencing research collaboration between the two nations, aiming to support the formulation of targeted research collaboration policies.

3.3.1. Model design

Scientific collaboration represents a form of cultural exchange between two countries. The gravity model has been employed in cross-cultural research to explain the extent of interaction and exchange between two nations or cultural entities [28,29]. This study utilizes the gravity model as the foundational framework, which is then mathematically transformed into a model appropriate for linear regression analysis. Additionally, regression analysis is conducted using paper citation data from the Web of Science (WoS) and the annual GDP data of the two countries to assess the impact of scientific research strength and economic development levels on bilateral scientific collaboration.

The most fundamental expression of a gravitational model in physics is:

$$F_{ij} = \mathrm{G}\frac{M_i M_j}{D_{ij}}. \tag{1}$$

In the formula, F_{ij} represents the gravitational force between objects *i* and *j*, *G* is a constant, M_i and M_j denote the masses of objects *i* and *j* respectively, and D_{ij} is the distance between objects *i* and *j*. The formula indicates that the gravitational force between two objects depends on their masses and is inversely proportional to the square of the distance between them. Introducing eq. (1) into the study of research collaboration between China and Malaysia, we construct a gravitational model for research collaboration between the two nations. This generalizable model can be extended for research collaboration between any two countries. Its mathematical expression is as follows:

$$C_{ys} = \mathrm{A}\frac{R_{ys}^{CHN} R_{ys}^{MY}}{D}. \tag{2}$$

In the formula, C_{ys} represents the quantity of research collaboration papers in different subjects *s* between China and Malaysia in the year *y*, *A* is a constant term, R_{ys}^{CHN} and R_{ys}^{MY} respectively denote the total number of research papers published by China (CHN) and Malaysia (MY) in different subjects *s* in year *y*. *D* is a constant representing the straight-line distance between the capitals of China and Malaysia. For computational convenience, eq. (2) is transformed into a logarithmic form as follows:

$$\ln C_{ys} = \alpha_0 + \alpha_1 \ln R_{ys}^{CHN} + \alpha_2 \ln R_{ys}^{MY} + \varepsilon_{ys}. \tag{3}$$

Economic development level is a significant factor influencing education and scientific research capacity. From a cross-cultural perspective, it is also partially associated with a nation's ability to shape international discourse. This study incorporates the economic development levels of the two countries as a key influencing factor in the model, quantified using GDP data. The revised model, which includes this new variable, is presented as:

$$\ln C_{ys} = \alpha_0 + \alpha_1 \ln R_{ys}^{CHN} + \alpha_2 \ln R_{ys}^{MY} + \alpha_3 G_y^{CHN} + \alpha_4 G_y^{MY} + \varepsilon_{ys}. \tag{4}$$

The dependent variable C_{ys} represents the quantity of research collaboration papers in different subjects s between China and Malaysia in the year *y*; R_{ys}^{CHN} and R_{ys}^{MY} respectively denote the total number of research papers published by China (CHN) and Malaysia (MY) in different subjects *s* in the year *y*. G_y^{CHN} and G_y^{MY} represent the per capita GDP of China and Malaysia, respectively, in the year *y*. *α* is an estimated parameter, and ε_{ys} is the random disturbance term.

3.3.2. Variable descriptions

The study includes two types of variables: (1) Dependent Variable: The dependent variable C represents the quantity of research collaboration papers between China and Malaysia (1999–2022), sourced from Web of Science. (2) Independent Variables: Independent variables include R, the total number of research papers published by China and Malaysia respectively (1999–2022, Web of Science), and per capita GDP of both countries, denoted as G, sourced from the World Bank.

3.3.3. Results

Basic Regression Equation Analysis:

Using the model (eq. 3), 120 qualified samples were analyzed in SPSS27. The regression results, shown in Table 1, indicate that both independent variables in the basic gravitational model (eq. 3) are highly significant, demonstrating its effectiveness in explaining research collaboration between China and Malaysia. Subsequently, the explanatory variables were expanded to include all additional factors in the extended model (eq. 4).

Table 1

Basic regression equation results

	Unst. Coef.	St. Coef.	t	Sig.	Adjusted R Square	F	Durbin-Watson
Const.	–27.553		–1.970	.051			
$\ln R_{ys}^{CHN}$	0.003	1.126	9.827	***	0.787	F=221.252 p < 0.001	2.219
$\ln R_{ys}^{MY}$	–0.016	–0.260	2.272	*			

****p <0.001; **p <0.01; *p <0.05*

Analysis of the Extended Regression Equation

Table 2 shows that all explanatory variables in the extended model (eq. 4) are significant. The F-value is 158.06 (p < 0.001), confirming the model's validity, with an Adjusted R Square of 0.841 indicating a strong fit. The Durbin-Watson value of 2.67 suggests no autocorrelation, validating the model's robustness. At a 95% confidence level, the number of research papers and GDP for both countries significantly affect collaborative academic publications. China's indicators show positive correlations, while Malaysia's are negative. Notably, the standardized coefficient of R for China's publications is 1.032, indicating the strongest influence in the model.

Table 2

Extended regression equation results

	Unst. Coef.	St. Coef.	t	Sig.	Adjusted R Square	F	Durbin-Watson
Const.	80.297		2.320	*			
$\ln R_{ys}^{CHN}$	.003	1.032	10.271	***	0.787	*F*=221.252 *p* < 0.001	2.219
$\ln R_{ys}^{MY}$	–.019	–.306	–2.941	**			
G_y^{CHN}	*A*	.759	6.150	***			
G_y^{MY}	*B*	–.586	–4.930	***			

A=0.000000000035412053891; B=-0.00000000143545060931;
****p <0.001; **p <0.01; *p <0.05*

4. RESULTS AND DISCUSSION

4.1. TRENDS IN ACADEMIC COLLABORATION AND IMPACT

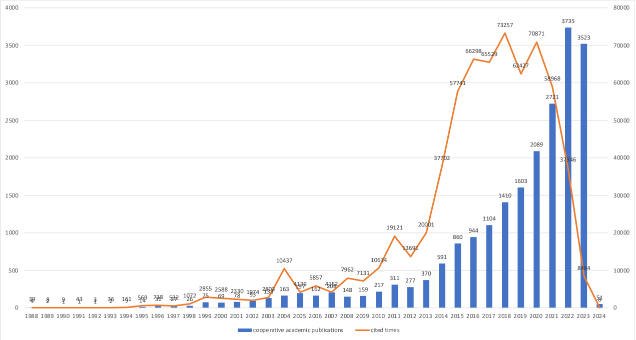


Fig. 4 – Cooperation academic publications and cited times in each year.

The quantity of co-authored academic publications (CAP) reflects the scale of academic collaboration between China and Malaysia, while citation counts indicate the impact of their cooperation. Figure 4 presents the CAP and citation counts analysis from 1988 onward, excluding incomplete data from 2023 and 2024. The findings reveal an overall exponential growth in CAP, which can be categorized into three stages:

- **Stage 1 (1988–1989) – Commencing:** In this stage, the number of CAP remained shallow, indicating minimal academic cooperation following the formal establishment of diplomatic relations between China and Malaysia [30].
- **Stage 2 (1999–2012) – Increasing:** In this stage, there was a noticeable increase in the number of CAP and a concurrent rise in citations. This suggests that academic cooperation between China and Malaysia started to develop towards a larger scale and higher quality. This aligns with the Joint Statement on the Future Bilateral Cooperation Framework between the Government of the People's Republic of China and the Government of Malaysia [31], which emphasized strengthening cooperation in science and technology, particularly in areas of mutual interest. Additionally, the China-Malaysia Joint Communiqué [32], mentioning more detailed areas of academic cooperation, corroborates the findings of this study.
- **Stage 3 (2013–present) – Rapid development:** In this stage, there has been a significant enhancement in the scope, scale, and impact of academic collaboration associated with establishing a comprehensive strategic partnership between China and Malaysia in 2013. Given that citation counts require a certain period to accumulate, the academic impact of China and Malaysia's academic collaboration is expected to continue to increase.

4.2. TRENDS IN RESEARCH AREAS

This study classified the data into five categories based on WOS's research field classifications officially provided: Arts & Humanities, Life Science & Biomedicine, Physical Sciences, Social Sciences, and Technology [15]. The results, as illustrated in Fig. 5, show distinct trends in each category:

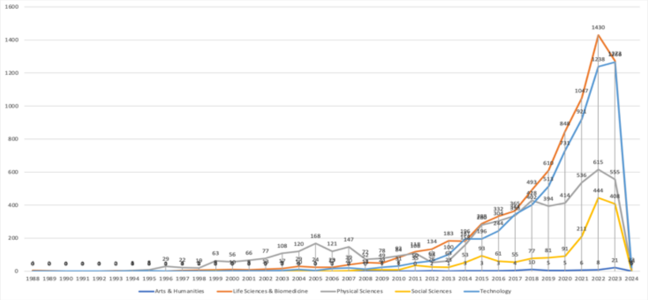


Fig. 5 – Research areas count in each year.

- **Physical Sciences:** The research trend in this category experienced fluctuations, with a minor peak in 2005 followed by a decline and a consistent rise from 2013, positioning it in the middle range among the five categories. This indicates that the field of physical science has always been a part of collaboration but has not been a highly popular area.
- **Life Science & Biomedicine and Technology:** The research trends in these categories are generally similar, exhibiting exponential growth from 2005 and significant potential for development. This suggests that since 2014, the main areas of collaboration have been concentrated in Life Science & Biomedicine and Technology. The trend indicates that these areas will continue to be the primary areas of cooperative research in the future, with an increasing proportion.
- **Social Sciences:** There was a noticeable increase in this category beginning in 2020, indicating that the social science field had not been a significant point of collaboration before this time. This might be related to the inherent characteristics of the research area.
- **Arts & Humanities:** This category has consistently maintained a very low level, suggesting minimal collaborative research between China and Malaysia in this field.

4.3. KEY TOPICS

This study performed cluster analysis on title data and word frequency analysis on abstracts within each cluster to identify key topics in China-Malaysia collaborative research. As shown in Fig. 3, clusters 0 and 3 are focused on physical sciences, while physical sciences dominate cluster 2 but also include life sciences and biomedicine. Clusters 4 and 5 encompass interdisciplinary topics.

Figure 1b indicates clear boundaries for clusters 0, 2, and 3, highlighting their distinctiveness with minimal overlap. In contrast, clusters 1, 4, and 5 show blurred boundaries due to dimensionality reduction (from 738 to 3), leading to information loss. This also suggests that these topics have intersectional areas rather than being entirely independent.

Based on the word frequency analysis, the key topics of China-Malaysia academic collaboration were identified as follows:

- **CLUSTER 0 – Molecular structure and design, as well as organics:** In cluster 0, terms frequently occurring, such as atom, molecule, structure, bond, and ring, are all chemistry-related terminologies. Most high-frequency words are general, which indicates that the chemistry research is extensive in scope and does not focus on research subjects with distinct characteristics of the two countries, such as the rubber industry. Besides, collaborative chemistry research appears to be highly specialized within chemistry itself, lacking integration of chemical discoveries with practical environmental or social science issues specific to both countries. This suggests an opportunity for future collaborative research that connects chemical findings more directly with these two countries' real-world environmental and social challenges.

- **CLUSTER 1 – Medical and life sciences:** In cluster 1, terms such as patient, treatment, health, and disease point to the field of medicine, while words like species, gene, and cell indicate a focus on life sciences. This suggests that medical research and life sciences are hot topics in research collaboration. Additionally, this cluster involves specific interdisciplinary research. For instance, research on COVID-19 in this cluster encompasses an interdisciplinary approach, including medicine, public health policy, and psychology. The results corroborate this aspect, indicating a comprehensive approach to addressing complex issues like COVID-19, which require insights from multiple scientific areas.
- **CLUSTER 2 & 3 – High-energy & particle physics:** The high-frequency words in these clusters include terms like quark, decay, mass, measure, and boson, which are related to high-energy physics or particle physics. These terms pertain to the study of the production, decay, and interactions of elementary particles, indicating that basic physics research, especially in the realm of particle and high-energy physics, is a focal point of collaboration between China and Malaysia in the areas of physics. Furthermore, since some physical research involves experiments that require advanced large-scale experimental equipment, these studies typically use facilities in China. For example, this paper used the Beijing Spectrometer (BESIII) Experiment equipment [33].
- **CLUSTER 4 – Engineering technology & material science:** In Cluster 4, terms such as property and structure indicate key topics in real estate and architectural engineering research. Meanwhile, words like material, use, and acid pertain to research key topics related to material development, encompassing areas such as food engineering, chemical engineering materials, and battery technology. This conclusion is also corroborated by the Fig 3.
- **CLUSTER 5 – Interdisciplinary research combining models, theories, and real-world issues:** In Cluster 5, terms like study, use, model, factor, and management point to research and management of practical problems using mathematical or computer models. For instance, the article on the predictive model to produce health foods researches the use of Long Short-Term Memory (LSTM) models to predict the production dynamics of powdered milk [34]. This study integrates the production of health foods with computer model predictions to address real-world issues in powdered milk production management. Combining the results and the cluster analysis, this cluster encompasses a combination of multiple disciplines and areas. For example, the article on greenhouse gas emissions and reforestation discusses the impact of different forest types on carbon emissions. It explores how to balance ecological restoration with the reduction of greenhouse gases. This represents an intersection of environmental science and social science.

4.4. ANALYSIS OF INFLUENCING FACTORS OF RESEARCH COLLABORATION

From the perspective of research capabilities, the quantity of publications from China significantly positively influences collaboration with Malaysia. This suggests a strong emphasis by China on collaboration with Malaysia in the research domain, consistent with prior research findings [4,5]. Conversely, Malaysia's publication quantity negatively impacts collaboration with China, indicating potential obstacles Malaysia faces in research collaboration with China, such as uneven distribution of research resources or differences in research priorities.

In terms of economic development, China's economic growth may stimulate increased research investment, reflected in the positive correlation between GDP and the quantity of collaborative publications with Malaysia. With economic growth, increased research funding may enhance the quantity and quality of research projects, thereby fostering more international collaborative publications. In contrast, the negative coefficient for Malaysia's GDP may suggest that, despite economic development, research resources may be more internally consumed, or Malaysia may prefer academic collaboration with other countries. Additionally, smaller economies might exhibit lower attractiveness or participation in research collaboration.

Moreover, in the linear regression’s results, China’s scientific research capacity and economic level exhibit a positive correlation with bilateral scientific collaboration, whereas Malaysia’s corresponding factors show a negative correlation. From a cross-cultural exchange perspective, these findings indicate the following:

Firstly, China assumes a dominant role in bilateral scientific collaboration. China actively expands its influence by investing substantial resources and advancing its scientific research capacity, serving as the primary driver of collaboration. This leadership role is also evident in China’s cultural exchanges and diplomatic activities with other ASEAN nations.

Secondly, Malaysia approaches bilateral collaboration conservatively. As a smaller economy, Malaysia faces significant constraints on resource allocation and tends to exercise greater caution in selecting collaboration partners and fields. Consequently, Malaysia’s economic growth or increased research output is more likely to focus on domestic collaborations or addressing local priorities rather than committing resources to international partnerships.

Thirdly, Malaysia is more inclined to collaborate with neighboring countries. Scientific collaboration is a form of cultural exchange, and cross-cultural interactions are influenced by the cultural distance between entities. As a multicultural nation where ethnic Chinese comprise only about one-third of the population, Malaysia may find it more natural to engage in scientific collaboration with neighboring Southeast Asian countries.

Lastly, China is better positioned to overcome cultural barriers. With robust economic support and high research output, China can mitigate some communication challenges posed by cultural differences, thereby fostering collaboration. In contrast, Malaysia, as a smaller economy with limited resources, may struggle to surmount these barriers. As a result, Malaysia's scientific collaborations are more likely to focus on countries with closer linguistic and cultural affinities, rather than prioritizing China as a partner.

The outcomes of research collaboration between the two countries are influenced by multifaceted interactions. For instance, economic growth may provide funding, but effective international collaboration depends on favorable policy environments and support for cultural communication. The modern relationship between China and Malaysia is built upon long-term economic cooperation and cultural exchange. Government agreements, such as research funding programs, have facilitated academic collaboration between the two nations. Cross-cultural exchanges play a critical role in promoting international academic cooperation. China's cultural and educational exchange programs, like Confucius Institutes, may enhance cultural understanding and language learning between China and Malaysia, thereby reducing collaboration barriers. Therefore, individual factors cannot fully explain the variations in the quantity of collaborative publications between China and Malaysia. The consideration of multifaceted interactions is essential.

5. **CONCLUSION**

This study employed cluster analysis and linear regression analysis to examine the collaborative dynamics between China and Malaysia from 1988 to the present. The findings indicate an overall exponential increase in collaborative publications between the two nations. Life sciences & biomedicine, and technology have emerged as the primary collaborative fields, showing a continuous upward trend. Conversely, collaboration in arts & humanities and social sciences, particularly before 2020, is relatively limited. The hot topics in collaboration concentrate on chemistry, medicine, physics, engineering, materials science, and interdisciplinary studies. China and Malaysia's research capabilities and economic development significantly influence collaborative research. Notably, the correlations are positive in China and harmful in Malaysia.

CREDIT AUTHORSHIP CONTRIBUTION STATEMENT

Guansu Wang: Core idea, analysis, and conclusions.
Zhihong Huang: Data collection, processing, and analysis.
Sameer Kumar: Quality supervision and manuscript oversight.

REFERENCES

1. J. Zou, C. Liu, G. Yin, Z. Tang, *Spatial patterns and economic effects of China's trade with countries along the Belt and Road*, Progress in Geography, **34,** ***5***, pp. 598–605 (2015).
2. C.B. Ngeow, *The five areas of connectivity between Malaysia and China: Challenges and opportunities*, The Belt and Road Initiative: ASEAN Countries' Perspectives, **8**, pp. 117–139 (2019).
3. M.N.M Akhir, L.C. Leong, H.M. Tahir, *Malaysia-China bilateral relations, 1974-2018*. WILAYAH: The International Journal of East Asian Studies, **7,** *1*, pp. 1–26 (2018)
4. J.L. Ding, L.Y. Yang, H.R. Sun, X.W. Liu, X.Y. Huang, T. Yue, L.Y. Chen., M.M. Zhu, F.Y. Chen, X.Z. Wang, *Bibliometric study on research collaboration among the Belt and Road areas and countries*, Bulletin of Chinese Academy of Sciences, **32,** *6*, pp. 626–636 (2017).
5. J.M. Zhou, Y. Huang, X.F. Wang, Y. Chen, Y. Fu, P.P. Ma, *Research on the research cooperation situation between China and the countries along the Belt and Road --- econometric analysis based on Web of Science*, Intelligence Engineering, **2,** *4*, pp. 69–79 (2016).
6. S. Kumar, Jan, J. Mohd., *Mapping research collaborations in the business and management field in Malaysia, 1980–2010*, Scientometrics, **97,** *3*, pp. 491–517 (2013).
7. Y.P. Ye, W.C. Ma, G.Y. Zhang, *Study on the current status of S&T cooperation between China and countries along the "Belt and Road" - a comparative analysis based on patents and papers*, Documentation, Information & Knowledge, *4*, pp. 60–68 (2016).
8. J.H. Chen, M.N. Xu, *Analysis of the situation and influence factors of China-ASEAN scientific research cooperation*, Journal of Information Resources Management, **10,** *2*, pp. 107–117 (2020).
9. M. Yu Cheng, K. Wah Hen, H. Piew Tan, K. Fai Fok, *Patterns of co-authorship and research collaboration in Malaysia*, Aslib Proceedings: New Information Perspectives, Emerald Group Publishing Limited, 2013.
10. H. Bukvova, *Studying research collaboration: a literature review*, All Sprouts Content, **10**, *3* (2010).
11. J. Davidson Frame, M.P. Carpenter, *International research collaboration*, Social studies of science, **9,** *4*, pp. 481–497 (1979).
12. T. Plotnikova, B. Rake, *Collaboration in pharmaceutical research: exploration of country-level determinants*, Scientometrics, **98,** *2*, pp. 1173–1202 (2014).
13. N. Song, X. He, Y. Kuang, *Research hotspots and trends analysis of user experience: Knowledge maps visualization and theoretical framework construction*, Frontiers in Psychology, **13** (2022).
14. L. Šubelj, N.J. Van Eck, L. Waltman, *Clustering scientific publications based on citation relations: a systematic comparison of different methods*, PLOS ONE, **11,** *4* (2016).
15. Web of Science., *Research Areas.* http://webofscience.help.clarivate.com/en-us/Content/research-areas.html (2023).
16. S. Bird, E. Klein, E. Loper, *Natural language processing with Python: analyzing text with the natural language toolkit*, O'Reilly Media, Inc. (2009).
17. T. Mikolov, I. Sutskever, K. Chen, G.S. Corrado, J. Dean, *Distributed representations of words and phrases and their compositionality*, Advances in Neural Information Processing Systems, **26** (2013).
18. T. Mikolov, K. Chen, G. Corrado, J. Dean, *Efficient estimation of word representations in vector space*, arXiv. arXiv:1301.3781 (2013)
19. J. Ramos, *Using TF-IDF to determine word relevance in document queries*, Proceedings of the First Instructional Conference on Machine Learning, (2003).
20. T.B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, *Language models are few-shot learners*, Advances in Neural Information Processing Systems (NeurIPS), **33** (2020).
21. J. Devlin, M.-W Chang, K. Lee, K. Toutanova, *BERT: Pre-training of deep bidirectional transformers for language understanding*, arXiv, arXiv:1810.04805 (2019).
22. S. Selva Birunda, R. Kanniga Devi, *Review on word embedding techniques for text classification*, Innovative Data Communication Technologies and Applications, Springer Singapore (2021).
23. I. Beltagy, K. Lo, A. Cohan, *SciBERT: A pretrained language model for scientific text*, Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP) (2019).
24. X. Cai, S. Liu, L. Yang, Y. Lu, J. Zhao, D. Shen, T. Liu, *COVIDSum: A linguistically enriched SciBERT-based summarization model for COVID-19 scientific papers*, Journal of Biomedical Informatics, **127**, 103999 (2022).
25. A. Glazkova, *Identifying topics of scientific articles with BERT-based approaches and topic modeling*, Trends and Applications in Knowledge Discovery and Data Mining, Springer International Publishing (2021).
26. P. Lobanova, P. Bakhtin, Y. Sergienko, *Identifying and visualizing trends in science, technology, and innovation using SciBERT*, IEEE Transactions on Engineering Management, **71**, pp. 11898–11906 (2023).
27. R. Nainggolan, R. Perangin-angin, E. Simarmata, A.F. Tarigan, *Improved the performance of the K-Means cluster using the sum of squared error (SSE) optimized by using the elbow method*, Journal of Physics: Conference Series, **1361,** *1*, 012015 (2019).
28. A. Golovko, H. Sahin, *Analysis of international trade integration of Eurasian countries: gravity model approach*," Eurasian Economic Review, **11,** *3*, 519–548 (2021).
29. X.N. Zhang, W.W. Wang, R. Harris, G. Leckie, *Analysing inter-provincial urban migration flows in China: A new multilevel gravity model approach*, Migration Studies, **8**, *1*, 19–42 (2020).
30. Chinese and Malaysian Governments., *Joint Communiqué of the Government of the People's Republic of China and the Government of Malaysia* (1974).
31. Chinese and Malaysian Governments., *Joint statement on the future bilateral cooperation framework between the Government of the People's Republic of China and the Government of Malaysia* (1999).
32. Chinese and Malaysian Governments., *China-Malaysia joint communiqué* (2004).
33. S. Jia, X.L. Wang, C.P. Shen, C.Z. Yuan, I. Adachi, H. Aihara, K. Senyo, *Observation of $e+\ e-\longrightarrow \gamma\ \chi\ c\ 1$ and search for $e+\ e-\longrightarrow \gamma\ \chi\ c\ 0,\ \gamma\ \chi\ c\ 2$, and $\gamma\ \eta\ c$ at s near 10.6 GeV at Belle*, Physical Review D, **98,** *9*, 092015 (2018).
34. O.A. George, A. Putranto, A., J. Xiao, P.S. Olayiwola, X.D. Chen, J. Ogbemhe, T.J. Akinyemi, A. Kharaghani, *Deep neural network for generalizing and forecasting on-demand drying kinetics of droplet solutions*, Powder Technology, **403**, 117392 (2022).